AL/CF-TR-1993-0164

AD-A275 126

INFORMATION PRESENTATION FOR EXPERT SYSTEMS IN FUTURE FIGHTER AIRCRAFT (U)

ARMSTRONG LABORATORY

Mica R. Endsley
Esin O. Kiris

DEPARTMENT OF INDUSTRIAL ENGINEERING
TEXAS TECH UNIVERSITY
LUBBOCK, TEXAS 79409

3198 94-03121

SEPTEMBER 1993

FINAL REPORT FOR THE PERIOD DECEMBER 1992 - AUGUST 1993

Approved for public release; distribution is unlimited

94 1 31 171

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICES

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1993-0164

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This Technical Report is published as received and has not been edited by the Technical Editing Staff of the Armstrong Laboratory.

FOR THE COMMANDER

for **KENNETH R. BOFF**, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
Sep 1993

3. REPORT TYPE AND DATES COVERED
Final Dec 92 - Aug 93

4. TITLE AND SUBTITLE
Information Presentation for Expert Systems in Future Fighter Aircraft

5. FUNDING NUMBERS
C - F33616-89-C-0532
PE - 62202F
PR - 7184
TA - 10
WU - 44

6. AUTHOR(S)
Mica R. Endsley, PhD, P.E.
Esin O. Kiris

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Industrial Engineering
Texas Tech University
Lubbock TX 79409

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Crew Systems Directorate
Human Engineering Division
Human Systems Center
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/CF-TR-1993-0164

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This study explored the method for presentation of information in the context of a crew-aiding "intelligent associate." Forty-five male Air Force pilots and forty-five male and female students served as subjects in the evaluation of prototype display designs. Categorical information coding produced the shortest processing and response times for both groups of subjects. Confidence in decision making was shown to be affected by both coding schema and subject experience level.

14. SUBJECT TERMS
Situational awareness, information portrayal/coding, decision making, artificial intelligence

15. NUMBER OF PAGES
31

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

THIS PAGE LEFT BLANK INTENTIONALLY

PREFACE

This effort was conducted by Texas Tech University under purchase order 3548 of Air Force Contract Number F33615-89-C-0532 issued by Logicon Technical Services Incorporated, Daytron, Ohio. The Air Force Contract Monitor was Mr. Robert Linhart. Technical guidance for the effort was provided by Mr. Gilbert G. Kuperman, Crew Station Integration Branch, Human Engineering Division, Crew Systems Directorate, Armstrong Laboratory, Wright-Patterson Air Force Base, Ohio. The effort was carried out in support of Work Unit 7184 10 44, "Advanced Strategic Cockpit Engineering."

The authors would like to express their appreciation to Captain Judy Burk and Captain Kay at Reese Air Force who assisted in the logistics of the data collection. Special appreciation is also due to the many pilots assigned to Reese Air Force Base who volunteered their time and expertise in the conduct of this study.

SUMMARY

Expert systems have been promoted as a means of reducing workload and providing improved decision support to pilots in advanced future aircraft. In order for these systems to be utilized effectively, a means of providing the system's recommendations and information for assessing the quality of those recommendations must be provided in a manner that meets the stringent workload and time requirements of the cockpit. A research study was undertaken to determine interface guidelines for presenting information on expert system recommendations in this context. Four methods of presenting expert system confidence associated with recommendations were compared to each other and to a control condition in which no confidence information was presented. Significant differences between display conditions and between experts and novices were found in their use of system confidence information. Recommendations are presented for conveying real-time information on the reasoning processes of expert systems in future cockpits.

Table of Contents

SECTION		PAGE
1	INTRODUCTION	1
2	OBJECTIVE	7
3	METHODOLOGY	8
	EXPERIMENTAL DESIGN	8
	TASKS	8
	SUBJECTS	10
	HYPOTHESES	11
4	RESULTS	13
	ANALYSES OF RESPONSE TIME AND CONFIDENCE LEVEL	13
	ANALYSES OF OPTIONS SELECTED	18
5	DISCUSSION AND CONCLUSIONS	19
REFERENCES		22

List of Figures

FIGURE	TITLE	PAGE
1	An aircraft scenario	9
2	An automobile scenario	10
3	Subject's mean decision time for presentation type and task type interaction	16
4	Subject's mean decision time for presentation type and subject type interaction	17
5	Subject's mean decision time for task type and subject type interaction	17

List of Tables

TABLE	TITLE	PAGE
1	Student and pilot subjects for aircraft and automobile tasks	11
2	ANOVA of subject response time	14
3	ANOVA of subjective confidence level	15

THIS PAGE LEFT BLANK INTENTIONALLY

Section 1

INTRODUCTION

The potential advantages of the application of intelligent decision support in the cockpit have been well documented. These include the reduction of pilot overload (O'Shannon, 1986), and the ability to overcome human shortcomings such as channelized attention, spatial disorientation and cognitive overload (McNeese, 1987). In addition, the need to extend human capabilities has been cited, particularly as may be required with the increased speeds and sophisticated avionics systems in more advanced military combat flight environments (McNeese, 1987; Summers, 1986).

This decision support may take the form of task automation, up to and including the development of intelligent systems for performing higher-order decision tasks. Although it may be theoretically feasible to completely automate some of these tasks, for technical, practical and social reasons, a form of decision support in which the pilot retains ultimate control is usually proposed. In this type of scenario the system may collect, integrate, transform and display required information, generate recommended actions using internal rules, and even carry out those actions upon the command of the pilot. Although many implementations of this "intelligence" are possible, the most widely used systems fall into the category of expert or knowledge-based systems.

The ultimate success of these endeavors will largely depend upon a full exploration of human factors issues inherent in decision support system implementation in the cockpit. This includes the selection of functions to be automated, elicitation of decision information from pilots, development of effective function allocation schemes, and the creation of a human interface which meets the stringent demands of the dynamic flight environment (Endsley, 1987). The increased complexity accompanying these systems will place a particularly high emphasis on the need for a good user

interface, as has been documented in work with expert systems in a variety of arenas (Berry & Hart, 1991; Wexelblat, 1989).

A suitable interface will be necessary in order for pilots to adequately assess the information provided by the systems and integrate it with their own knowledge to formulate a desired course of action. Unless the interface allows this to occur easily and rapidly, the system may not be used to its potential and may even hinder performance rather than help it. Eggleston (1992) points out that "the value of the aiding ... depends on whether or not its mission impact exceeds the cost of using it". Aretz, Guardino, Porterfield and McClain (1986), for instance, found that the additional resources required to request advice from an expert system were sufficient to decrease overall mission performance in a simulated flight task. The automatic presentation of this information resulted in a significant improvement over presentation on request, however, and resulted in mission performance above that of a control (no advice) condition. In order for these systems to provide the desired benefits in workload reduction, they must be well integrated with the tasks of the pilot and must not demand more from the pilot for their use than that incurred without them. Achieving this goal depends on the design of the interface between the pilot and the decision support system.

Some interface guidelines can be derived from established human-computer interface design principles. Expert systems, however, often involve additional human interface issues. In order that systems can be used effectively, users must have an adequate conceptual model of what the system does, and be able to interact with it. This requires their being able to assess whether or not the system could be used to help with a particular problem, to be able to input any data correctly, to assimilate any output, and to combine system advice with their own knowledge about the problem in order to reach a conclusion. (Berry & Hart, 1991).

Effective decision support requires that pilots be able to quickly determine the system's recommendations for a particular action and derive sufficient information for assessing the goodness

of that recommendation. As the ability of users to adequately weigh system recommendations and integrate this knowledge with their own depends on an assessment of the quality of those recommendations, particular emphasis must be placed on providing efficient and timely transfer of information on the decision processes used to arrive at them. Two primary issues are inherent in developing this level of understanding: the possession of a good mental model of how the system operates and the ability to determine on a case by case basis why a particular recommendation was generated.

The need for the user to have a good model of the system has been widely discussed. In order for pilots to achieve trust in the system — to determine when to trust the system (and when not to) — they will need to understand why the system makes the decisions it does and what factors it does and does not consider. Klein and Calderwood (1986) observe "in the absence of trust, it is not clear what evidence can help non-experts evaluate the quality of the answers they are receiving".

Hall (1985) found that subjects who had a good mental model of an expert system (as generated by a detailed description of the rules, inference networks and backward chaining procedures used by the system) needed fewer queries to determine why the system generated its diagnosis and reported greater subjective understanding of the system and ease of use than did those with only cursory information on which to form a mental model. Wexelblat (1989) recommends encouraging what-if experimentation and logging errors for user review to help users develop a good mental model of an expert system.

Even with a good model of the system, however, users may need more information on why a particular recommendation was made. In the cockpit, new considerations associated with communicating the reasoning process of the system may be present that are not present in static ground-based systems. The how and why facilities typically provided for expert systems are probably far too cumbersome for time critical flight tasks. Very little has been done to provide this

type of capability in a system with the stringent decision time restrictions of the cockpit. A more direct form of information presentation may be necessary to convey the decision process of the expert system.

Investigating this issue, this paper presents research on the use of expert systems for supporting decisions under uncertainty in future aircraft systems. In general, the output of an expert system is not deterministic, but rather probabilistic. That is, it makes a decision by selecting the option that has the highest probability of being correct according to internal rules that apply to the situation. The best format for providing the pilot with information on this process needs to be determined, however.

This matter will be particularly important with cockpit expert systems which use direct sensor data as input. In the past, the confidence level of data, largely determined by the sensor source and its foibles, was obvious and key to the pilot's decision processes. The expert system will be likely to obscure this type of information by automatically obtaining the data and processing and fusing it with other data to arrive at its decisions. Thus, the pilot will have even less information on how much trust or confidence to place on a particular decision than without the system's assistance, unless a means of compensation can be found.

Two approaches for dealing with this issue in the aircraft cockpit are described by Emerson, Reising and Britten-Austin (1987). They discuss the use of uncertain data by the Electronic Crewmember (EC) and describe two possible approaches for dealing with probabilistic data: (1) The EC could represent uncertainty to the pilot using probability "tags," thus allowing the pilot to resolve the uncertainty while maintaining awareness of it, or (2) the EC could resolve the uncertainty itself using preprogrammed rules, thus reducing decision workload on the pilot. A major danger in removing the pilot from the decision process with the later option is that information about the situation and decision options can be lost, resulting in a loss of information that may be important for

building situation awareness and/or forming decisions later in the flight. In light of this, many pilots have indicated that they prefer to have aiding systems provide them with an indication of the probabilities associated with various options (i.e. the system's confidence level in its recommendations), leaving the pilot with the ultimate decision and control.

The implementation of probability information in a dynamic environment such as the cockpit is not that straight-forward, however. Humans in general are rather poor at dealing with statistical data (Wickens, 1992). Kidd and Cooper (1985) comment on the degree to which users were able to cope with probability information associated with expert system recommendations for a fault diagnosis operation. They observed that numerical probabilities were not easily understood by users. Translating these numbers into categories, while potentially reducing difficulty somewhat, was believed to reduce the amount of information provided to the user about the knowledge used to generate the system recommendations. They conclude that probabilities displayed to users be evaluated on the basis of appropriate coarseness of scale, performance sensitivity, user intelligibility and necessity.

Selcon (1990) found that presenting the probabilities associated with different options from a decision support expert system improved subject decision time and confidence ratings only when the probabilities were clearly different. When the probabilities were more similar, leading to some ambiguity as to what to do, subject decision time was slower than if no probability information had been presented at all. This indicates that more research is needed to determine the feasibility and desirability of displaying probabilities associated with decision options. Klein and Calderwood (1986) speculate on the problems associated with providing such abstract data to decision makers under pressure. They believe that the use of analogies and prototypes may have greater acceptance than probabilistic estimates. It is unclear, however, how to implement this recommendation with many types of systems.

In addition, the effect of pilot experience needs to be considered in determining the expert system interface. It would be expected that expert system advice would be most helpful to those with less experience to draw upon to form decisions, as hypothesized by Morris, Rouse and Frey (1984). A study of decision aiding by Aretz, Guardino, Porterfield and McClain (1986) did not find this, however. The authors reported that as information processing requirements in the simulated flight task of their study were quite high, even pilots with a high level of expertise benefited from expert system information. It may be, however, that the information presentation needs of pilots with greater expertise may differ widely from that of more novice pilots. Therefore, if decision support systems are to be used by pilots with differing levels of experience, the unique information requirements of these groups should be considered.

Section 2

OBJECTIVE

The overall objective of this research was to determine a pilot compatible method for presenting confidence level information associated with recommendations by an expert system. It is hypothesized that the manner of presentation of confidence information will directly impact the utilization of that information (in terms of processing time and utility) and thus its effectiveness in supporting the pilot. It is furthermore hypothesized that this utilization will be affected by the level of expertise of system users. Specifically, it is expected that users with little expertise in an area will be more reliant on the recommendations of an expert system than users with more expertise and this difference will be reflected in their degree of compliance with expert system recommendations and time to make a decision.

Section 3

METHODOLOGY

EXPERIMENTAL DESIGN

The experiment was constructed as a between subjects design. The three independent variables (factors) included:

Method of presentation used to convey information about the expert system's confidence level concerning its recommendations (a) digital (e.g. 75 %), (b) categorical (high, medium or low), (c) analog bar (thermometers), (d) ranks (1, 2 or 3), and (e) no information (control);

Task type (a) automobile task, and (b) aircraft task; and

Subject type (a) students, and (b) pilots.

The dependent measures for each subject's performance were time to make a decision, the decision selected, and subjective confidence about the correctness of decisions made. Response time for each scenario was measured to the nearest thousandth of a second by the computer's clock. The confidence level was provided as a subjective estimate, measured from 1 (low) to 10 (high).

TASKS

Two tasks were created using Hypercard software running on a Macintosh computer: An aircraft task and an automobile task. The subject's task in both cases was to observe the presented scenarios and decide on one of the three possible actions as quickly as possible. The aircraft task was presented first followed by the automobile task.

Aircraft Task. Ten aircraft scenarios were created which provided a static picture of a cockpit situation awareness display. An example scenario is shown in Figure 1. An expert system was simulated for advising subjects on the best action to take, based on information elicited from ten experienced fighter pilots. Three options were shown on the right side of the static picture along with the system's assigned probability for each option. The subject was instructed to click on the button that corresponded to his/her choice using the mouse on the computer. After selecting an option, the next scenario was presented automatically. At the end of all ten scenarios, the subject's confidence in his/her decisions was elicited on a ten point scale.

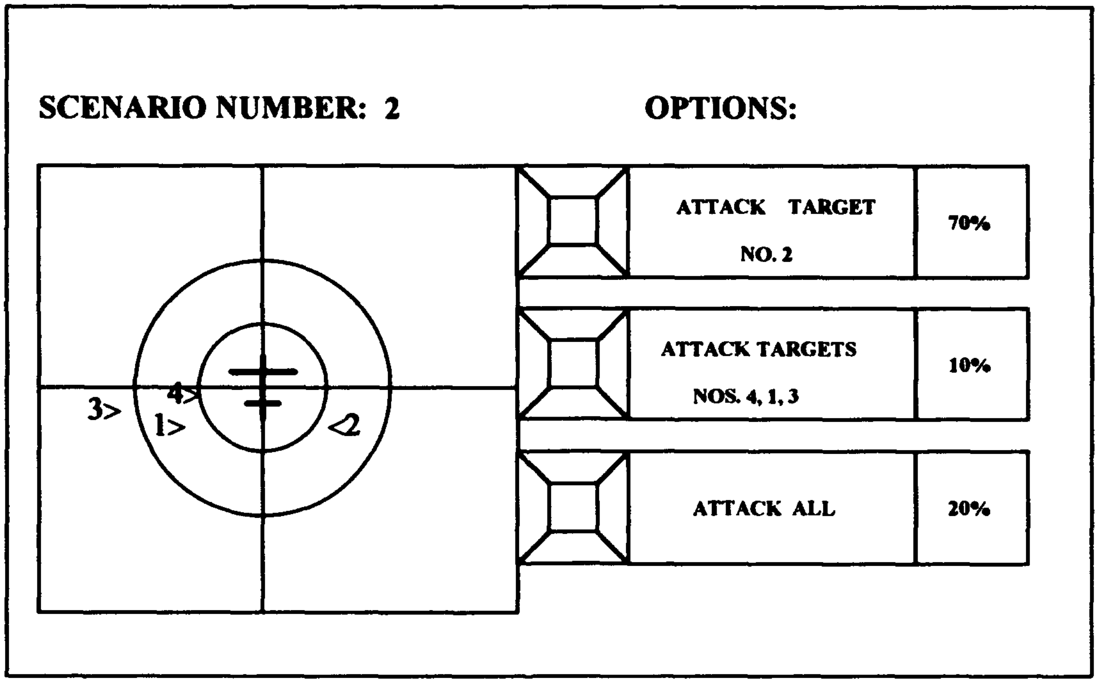

Figure 1. An aircraft scenario

Automobile Task. An automobile navigation task was created which depicted a real world driving situation (adapted from Selcon, 1990). For each of six scenarios a paragraph of text describing a decision task was presented. An example scenario is shown in Figure 2. An expert system was simulated which provided three decision options and confidence information regarding each. After reading the problem description, the subjects called up a list of three decision options

and assigned probabilities which were displayed under the problem paragraph. After subjects selected an option, the next scenario was presented. Subjects' confidence in their decisions was elicited at the end of the task on a ten-point scale.

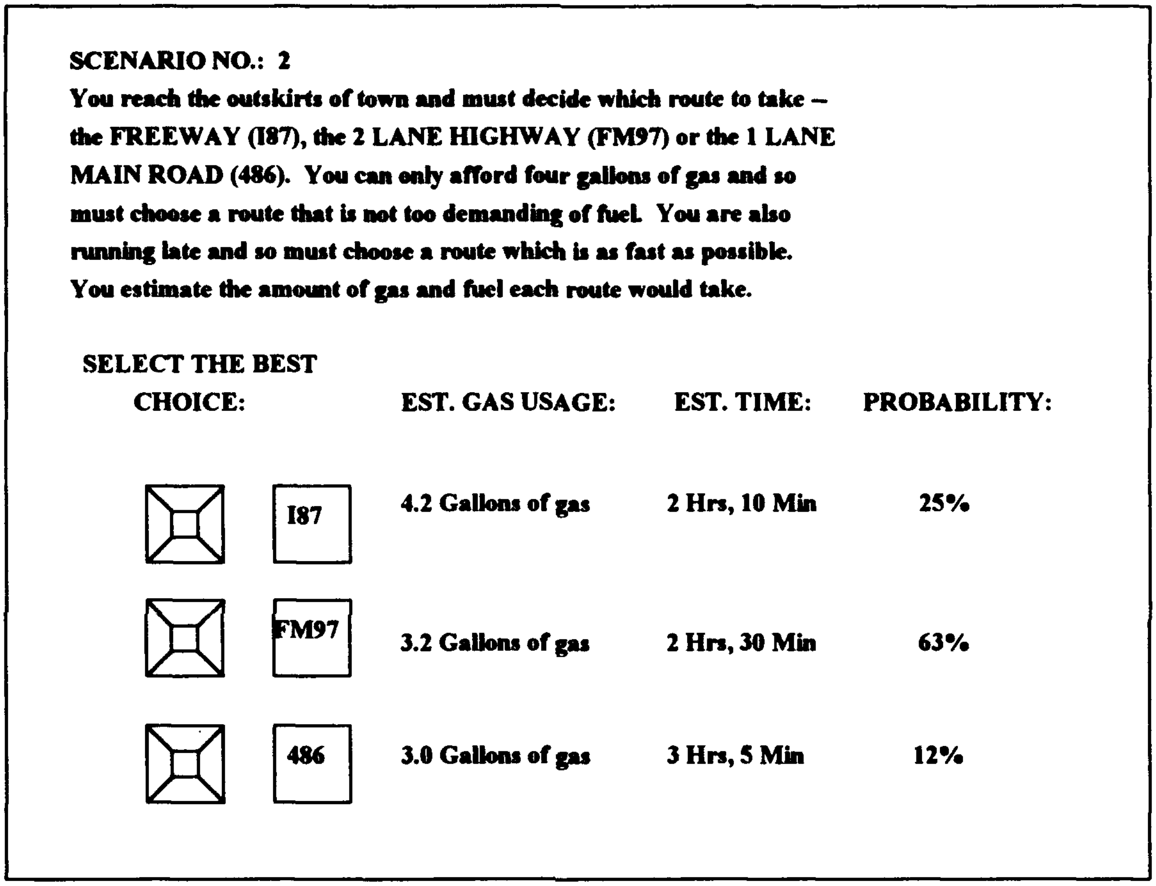

Figure 2. An automobile scenario

SUBJECTS

Two types of subjects participated in both tasks, as shown in Table 1. First, 45 available undergraduate and graduate students (40 male, 5 female) at Texas Tech University were recruited on a voluntary basis. Student subjects' mean age was 25.6 years with a variance of 14.1 years. This population was believed to possess a level of expertise which is representative of the general

population on the driving task, but could be classified as novice on the aircraft task as they had no prior experience in flying aircraft or performing tactical flight tasks.

In addition, 45 male U. S. Air Force pilots participated on a voluntary basis. This group was believed to have a high level of expertise on both tasks. Pilot subjects' mean age was 33.9 years and the variance was 57.9 years. The pilots were highly experienced with 2092 mean flight hours (range 570 to 5000) and 10.1 mean years of flying (range 3 to 22). Of these, 75% were trained in tactical aircraft and 33% reported combat experience.

Table 1. Student and pilot subjects for aircraft and automobile tasks.

	Aircraft	**Automobile**
Students	Non-expert	Expert
Pilots	Expert	Expert

HYPOTHESES

First, it was hypothesized that each of the methods of presentation (digital, categorical, analog and rank) which was used to convey information about the expert system's confidence level would reduce decision making time as compared to no information (control), would increase subjective confidence as compared to no information (control) and would be different from each other in their effect on decision time and subjective confidence.

A comparison of the effect of presentation type between the aircraft and automobile tasks for the student subjects should also provide an indication of the effects of expertise on requirements for the presentation of expert system recommendations. This comparison was confounded, however by other inherent differences between the two tasks. (The automobile task is verbal and analytical by nature and the aircraft task is pictorial and holistic in nature.) For this reason, performance by the

student (non-expert) group will be compared to the behavior of pilots (experts) who should possess expertise on both tasks. Any differences observed between task types can then be attributed to true differences in the expertise level of the subject population or to other differences between two tasks.

It was hypothesized that the two tasks would induce differing levels of dependence on the expert system. The automobile task was predicted to produce less reliance on the expert system as subjects could figure out the scenarios unaided. The aircraft task was predicted to produce more reliance on the expert system for the student subjects as they had no training or experience to draw upon to make decisions, while the pilot subjects were considered to possess expertise in this task and would be expected to have less reliance. Therefore, it was hypothesized that student subjects will be more likely than pilot subjects to choose the number one choice recommended by the expert system in the aircraft task (when this information is presented) as they will be relying on the expert system, but will be equally likely to choose the number one choice in the automobile task where the two groups will be equally reliant.

Section 4

RESULTS

Response time data for 15 scenarios out of the total 1440 scenarios administered were omitted due to very long response times corresponding with distractions during data collection, or very short response times indicating data entry errors. Three omissions of confidence level data occurred (out of a total of 180 solicitations) due to lack of data entry by subjects. The data were analyzed as a three factor experiment: (1) the method of presentation (digital, categorical, analog, ranks, no information); (2) task type (aircraft and automobile tasks); and (3) subject type (student and pilot subjects).

ANALYSES OF RESPONSE TIME AND CONFIDENCE LEVEL

Results of an ANOVA for response time showed a significant effect ($\alpha < .01$) of task, presentation, subject type, a task by presentation interaction, a subject type by presentation interaction, a subject type by task interaction, and a three-way interaction between task, subject type and presentation. Results of an ANOVA for confidence level showed a significant effect ($\alpha < .05$) of task type but not for presentation type, subject type, task by presentation interaction, subject type by presentation interaction, or the three-way interaction of task, subject type and presentation. Results of the analysis of variance on decision time and confidence level are shown in Tables 2 and 3 respectively. A Tukey pairwise test at the $\alpha = .05$ level was conducted to investigate each of the significant main effects.

Presentation type. There was a significant effect of presentation type on response time, $F(4, 1405) = 13.99$, $p < .001$. Categorical information presentation (high, medium or low) had a lower mean decision making time (10.0 s) as compared to all other conditions. Although digital and analog conditions had slightly higher mean decision times (14.7 s and 14.0 s), they were not significantly

different than the control condition (12.9 s). The digital condition, however, had a significantly higher mean decision time than the rank condition (12.3 s).

Table 2. ANOVA of subject response time.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F_0	P
Presentation type	4	3698.281	924.570	13.990	0.000
Task Type	1	8104.812	8104.812	122.640	0.000
Subject Type	1	3839.561	3839.561	58.099	0.000
Presentation*Task	4	954.160	238.540	3.610	0.006
Presentation *Subject Type	4	1058.078	264.519	4.003	0.003
Task*Subject Type	1	2917.272	2917.272	44.144	0.000
Presentation*Task *Subject Type	4	901.404	225.351	3.410	0.009
ERROR	1405	92850.843	66.086		

Task type. There was a significant main effect of task type for both response time, F (1, 1405) = 122.64, p <.001, and confidence level, F (1, 157) = 4.327, p <.05. Subjects' average response time in the aircraft task (10.9 s) was significantly faster than their average response time in the automobile task (15.8 s). Subjects' average confidence level in the aircraft task (7.1), however, was less than in the automobile task (7.6).

Subject type. There was also a significant main effect of subject type on response time, F (1, 1405) = 58.10, p <.001. Pilot subjects (11.4 s) were significantly faster than student subjects (14.1 s) in mean decision making time.

Table 3. ANOVA of subjective confidence level.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F_0	P
Presentation type	4	7.413	1.853	0.653	0.625
Task Type	1	12.272	12.272	4.327	0.039
Subject Type	1	9.083	9.083	3.203	0.075
Presentation*Task	4	10.621	2.653	0.936	0.445
Present.*Subject Type	4	24.83	6.208	2.189	0.073
Task*Subject Type	1	0.486	0.486	0.171	0.680
Presentation*Task *Subject Type	4	5.933	1.483	0.523	0.719
ERROR	157	445.233	2.836		

Presentation Type and Task Type Interaction. The interaction effect between presentation type and task type was significant for response time, F (4, 1405) = 3.61, p =.006. Subjects' average response time is shown in Figure 3. In general, the trends were similar across both tasks, however, average decision time in the control condition was higher in the automobile task than in the aircraft task. The digital condition had significantly increased decision making time (as compared to the control condition) in the aircraft task, but not in the automobile task. The categorical condition did not have significantly reduced decision time in the aircraft task, but did in the automobile task. The analog condition had significantly increased decision time in the aircraft task, but not in the

automobile task. The rank condition did not significantly change the decision time in the aircraft task, but did reduce the decision time in the automobile task.

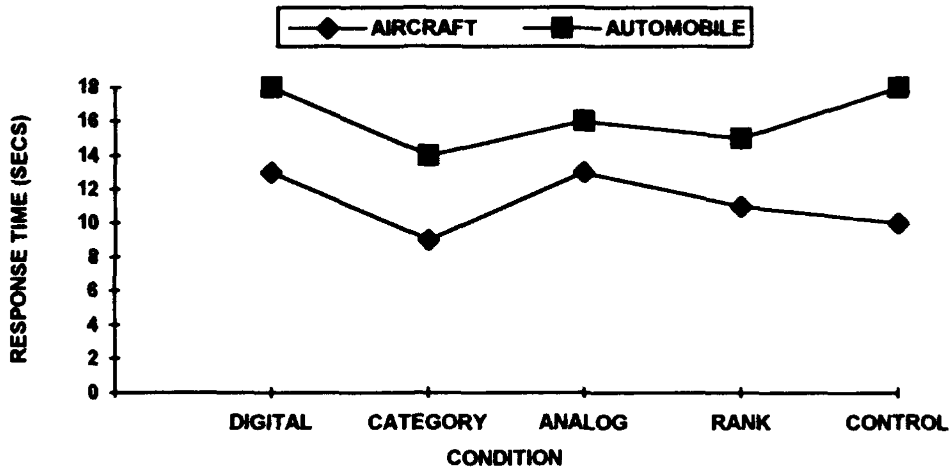

Figure 3. Subject's mean decision time for presentation type and task type interaction

Presentation Type and Subject Type Interaction. The interaction effect between presentation type and subject type was significant for response time, $\underline{F}$ (4, 1405) = 4.00, p = 0.003. Subjects' average response time is shown in Figure 4. In general, a similar trend was apparent across the presentation conditions for both groups, however these differences were only significant in some cases. Digital presentation did not have significantly increased decision making time (as compared to the control condition) for either students or pilots. The categorical condition had significantly reduced decision time for the students, but not for the pilots. Neither the analog nor the rank condition had significantly increased decision time for either students or pilots.

Task Type and Subject Type Interaction. The interaction effect between task type and subject type was significant for response time, $\underline{F}$ (1, 1405) = 44.14, p <.001. Subjects' average response time is shown in Figure 5. Pilots were significantly faster than students in the automobile

task, but not significantly faster than the students in the aircraft task. Overall, students were significantly slower for the automobile task compared to all other conditions.

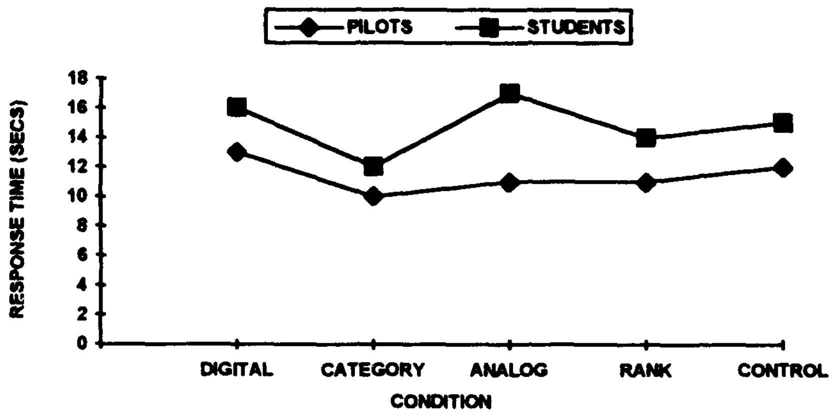

Figure 4. Subject's mean decision time for presentation tyne and subject type interaction

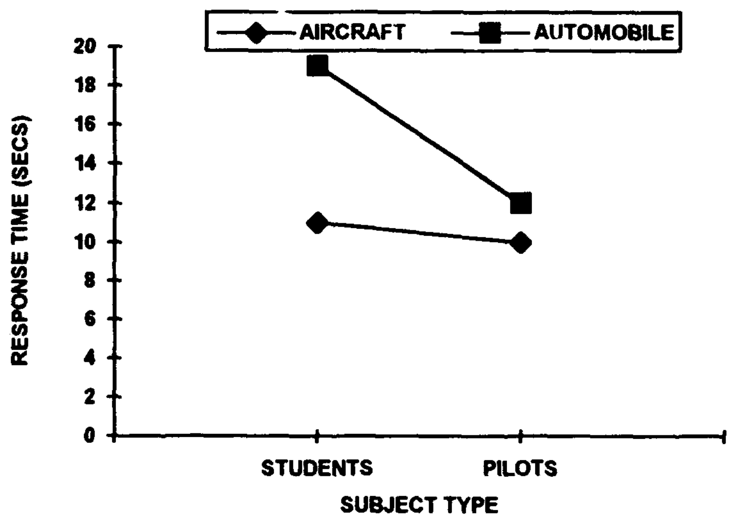

Figure 5. Subject's mean decision time for task type and subject type interaction

Three-Way Interaction. The interaction between task type, subject type and presentation type was significant for response time, $F(4, 1405) = 3.41$, $p = 0.009$. Pilot and student subjects were fastest with categorical information presentation (high, medium or low) on both tasks. The

lowest mean response time (8.7 s) was recorded by pilot subjects using the categorical presentation on the aircraft task. The highest mean response time (23.2 s) was recorded by the students using the digital presentation on the automobile task.

ANALYSES OF OPTIONS SELECTED

Overall, subjects were not more likely to have made different decisions in the different presentation conditions based on Chi-square tests at the $\alpha = 0.05$ level. Students and pilots as a group, however, were significantly different in their tendency to select the optimal option, $\chi^2 = 25.173$, $\alpha = 0.05$. Subjects also were significantly different in their likelihood of selecting the optimal option in the aircraft task as compared to the automobile task, $\chi^2 = 78.21$, $\alpha = 0.05$. In the aircraft task, students selected non-optimal alternatives 42.0% of the time, and pilots selected non-optimal alternatives 23.0% of the time. In the automobile task, students selected non-optimal alternatives 8.8% of the time, and pilots selected non-optimal alternatives 4.6% of the time.

Section 5

DISCUSSION AND CONCLUSIONS

Although it was hypothesized that the presentation of system confidence probabilities in the four formats (digital, categorical, analog and rank) would reduce decision making time as compared to no information (control), this does not appear to have been true for all conditions. Probability information from an expert system presented in a categorical format (high, medium or low) resulted in the quickest processing and response time for both novices and experts. Although it provided greater detail, the significantly greater amount of time required to process information with the digital format would indicate that this presentation form should be avoided. Analog presentation also increased decision making time, but only for the aircraft task and only for novices. Rank information did not appear to be significantly different from presenting no information across all conditions. Subjects' confidence in their decisions was not significantly impacted by the presence of system probability information in any of its presentation formats.

It is very interesting that across most presentation formats the expert system probability information impacted time to respond, even though subjects did not select the "best" alternative recommended by the expert system with greater frequency than when this information was not presented. Subjects appear to have been using the expert system information indirectly in conjunction with their own reasoning processes. With certain forms of presentation this extra processing actually adds to the decision making time, while in others it appears to reduce decision time somewhat. The greater detail provided by the digital and analog conditions appeared to slow down the novices quite a bit, but did not pose as great a problem for the experts.

Although it was hypothesized that student subjects (novices) would be reliant on the system's recommendations more than pilot subjects (experts), and in the aircraft task more than in the

automobile task, they did not appear to select the best response more often in any of the conditions where expert system information was presented. Student subjects were just as likely to pick low probability alternatives when the probabilities were shown as when they were not. This was true even for the aircraft scenarios, although student subjects had no expertise on which to base their decisions.

It was hypothesized that pilots possessed expertise on both tasks, as compared to students. This was confirmed, as pilot subjects were faster than student subjects. Further investigation revealed that they were only faster on the automobile task, however. The greater mean age and driving experience of the pilot subjects may have made them faster at the automobile task as compared to the student subjects.

For the aircraft task, on which the novices were expected to be slower, novices and experts had almost identical response times. It is believed that the novices may have been simply guessing and this resulted in a fairly fast response time. This is confirmed by the finding that novices were almost twice as likely as experts to choose a non-optimal alternative in the aircraft task, with a frequency which is relatively high (42%). They did not appear to take advantage of the expert system information, even when they had no other information on how to perform the task. There are several possibilities as to why this may have occurred. It is possible that the student subjects either did not trust the expert system or they did not care about the outcome associated with the task, whereas the pilots may have taken it more seriously.

Overall, these results call into question whether presentation of probability information is advisable. Before including such information, the features of a task and the skills of users should be identified. If speed is important, then a categorical form of presentation would be advised and digital and analog forms of presentation should be avoided, particularly for novices. However, the results of this study indicate that it may be advisable to pursue an information presentation strategy

which does not rely on probability information, given the lack of improvement in decision making in all forms of presentation.

REFERENCES

Aretz, A., Guardino, A., Porterfield, T., & McClain, J. (1986). Expert system advice: How should it be given? In *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 652-656). Santa Monica, CA: Human Factors Society.

Berry, D., & Hart, A. (1991). User interface standards for expert systems: Are they appropriate? *Expert Systems with Applications*, 2, 245-250.

Eggleston, R. G. (1992). Cognitive interface considerations for intelligent cockpits. In *Proceedings of the NATO/AGARD Symposium on Combat Automation for Airborne Weapon Systems: Man/machine Interface Trends and Technologies*. Edinburgh, Scotland: NATO - AGARD.

Emerson, T. J., Reising, J. M., & Britten-Austin, H. G. (1987). Workload and situation awareness in future aircraft. In *Recent advances in cockpit aids for military aircraft*. London: RAES.

Endsley, M. (1987). The application of human factors to the development of expert systems for advanced cockpits. In *Proceedings of the Human Factors Society 31st Annual Meeting* (pp. 1388-1392). Santa Monica, CA: Human Factors Society.

Hall, R. B. (1985). *Mental models and problem solving with a knowledge-based expert system* (85-108). McLean, VA: PAR Technology Corporation.

Kidd, A. L., & Cooper, M. B. (1985). Man machine interface issues in the construction and use of an expert system. *International Journal of Man Machine Studies*, 22, 91-102.

Klein, G., & Calderwood, R. (1986). Human factors considerations for expert systems. In *Proceedings of the National Aerospace and Electronics Conference (NAECON)* (pp. 921-925). New York: IEEE.

McNeese, M. D. (1987). Humane intelligence: A human factors perspective for developing intelligent cockpits. In *Proceedings of the National Aerospace and Electronics Conference (NAECON)* (pp. 941-948). New York: IEEE.

Morris, N. M., Rouse, W. B., & Frey, P. R. (1984). *Adaptive aiding for human-computer control: Conceptual model and experimental approach* (AFAMRL-TR-84-072). Wright-Patterson AFB, OH: Air Force Aerospace Medical Research Laboratory.

O'Shannon, D. (1986). Expert systems in EW. *DS & E*, March, 19-25.

Selcon, S. J. (1990). Decision support in the cockpit: Probably a good thing? In *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 46-50). Santa Monica, CA: Human Factors Society.

Summers, P. I. (1986). Cockpits for 2010 and beyond. *IEEE AES Magazine*, February, 17-20.

Wexelblat, R. L. (1989). On interface requirements for expert systems. *AI Magazine*, 66-78.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: Harper Collins.

152153

U.S.G.P.O.:1994–550-057/81051